Q1

February 19, 2000, all of which were published in English.

(3) On page 3, delete the first full paragraph and substitute the following paragraph:

In order to identify genes encoding PUFA elongases, it is necessary to study systems in which the synthesis of PUFAs is well documented; a good example of this is the model animal system *C. elegans*, a small free-living worm (Tanaka *et al.*, (1996), *Lipids* 31, 1173-78). *C. elegans*, like most other animals, and in contrast to higher plants, synthesizes PUFAs such as arachidonic acid (AA; 20:4 $\Delta^{5, 8, 11, 14}$) as precursors to a class of molecules known as the eicosanoids, which in turn serve as precursors for compounds such as prostaglandins and leucotriens (Horrobin, (1990), *Review in Contemp Pharmacotherapy*, 1:1-45). The presence of AA and other long chain polyunsaturated fatty acids in *C. elegans* is well documented (Tanaka *et al.*, (1996), *Lipids* 1, 1173-1178). The complete sequence of the nematode's genome is now publicly available (*The C. elegans cosortium, 1998, Science 282, 2012-2018.* See the database at the website identified with the URL file type, www host server, domain name sanger.ac.uk and following the path from "Projects" to "C_elegans" to "blast_server.shtml".

Q2

(4) On page 7, delete the paragraph beginning on line 18 and extending to page 8, line 21 and replace it with the following paragraph:

Initially the *C. elegans* databases were searched for any sequences which showed low levels of homology to yeast ELO genes (*ELO2* and *ELO3*) using the TBLASTN programme. A similar search was carried out using short (20 to 50 amino acid) stretches of ELO genes which were conserved amongst the three ELO polypeptide sequences. *C. elegans* sequences which were identified by this method were then used themselves as search probes, to identify any related *C. elegans* genes which the initial search with the yeast sequences failed to identify. This was necessary because the level of homology between the yeast ELO genes and <u>any</u> worm genes is always low (see BLAST scores later). To allow for a more sensitive search of worm sequences, a novel approach was adopted to circumvent the major drawback with searches using the BLAST programmes, namely that the search string (i.e. the input search motif) must be longer than 15 characters for the algorithm to work. Thus, if it was desired to search for a short motif (like a

Q3

histidine box), then the BLAST programme would not be capable of doing this. A complete list of all the predicted ORFs present in the *C. elegans* genome exists as a database called Wormpep, which is freely available from the Sanger WWW site identified with the URL address http file type, www host server, domain name sanger.ac.uk and following the path from "Projects" to "C_elegans" to "webace_front_end.shtml". The latest version of Wormpep was down loaded to the hard disc of a Pentium PC, and re-formatted as a Microsoft Word6 document, resulting in a document of about 3,500 pages. This was then searched using the "Search & Replace" function of Word6, which also allows for the introduction of "wildcard" characters into the search motif. So, for example, it is possible to search both for the short text string HPGG, which would identify any predicted worm ORF present in the Wormpep 3,500 page document containing this motif, or alternatively search with HPGX (where X is a wild card character). Clearly, such (manual) searches of a 3,500 page document are extremely time-consuming and demanding, also requiring visual inspection of each and every identified ORF. For example, searching with a motif such as HXXHH identifies in excess of 300 different ORFs. However, by using a number of different short search strings (as outlined below), and combining these with other methods for identifying putative elongase enzymes, a number of candidate ORFs have been identified.

(5) On page 8, delete the paragraph beginning at line 23 to and extending to page 9, line 3 and substitute the following paragraph:

As a negative control, to demonstrate that the FAE1 gene sequence was unlikely to provide a useful search sequence in the identification of *C. elegans* sequences encoding for PUFA elongases, the GenBank databases identified with the URL address http file type, www host server, domain name "ncbi.nlm.nih.gov" and following the path from Web to Search to index.html were searched using the *Arabidopsis* FAE1 polypeptide sequence ot identify related genes or expressed sequence transcripts (ESTs). GenBank is the NIH genetic sequence database, an annotated collection of all publicly available DNA sequences (*Nucleic Acid Research* (1998) 26, 1-7). There are approximately 2,162,000,000 bases in 3,044,000 sequence records as of December 1998. The search was carried out using the BLAST2 (Basic Local Alignment Search Tool) algorithm (Altschul *et al.*, (1990) *J Mol Biol* 215, 403, 410). Although a number of plant

ORFs and ESTs were reported as being related, o animal sequences were identified by this search, confirming the observation that FAE1 was unlikely to be a suitable candidate as a search template for PUFA elongases.

(6) On page 9, delete the paragraph beginning on line 5 and substitute the following paragraph:

Using the three yeast fatty acid elongase sequences (ELO 1, 2, 3) as probes, a number of putative ORFs in the DNA of *C. elegans*-derived cosmid sequences which form the *C. elegans* genomic sequence database was identified. Moreover, an extensive and time-consuming search of a downloaded copy of the WormPep database identified with the URL address ftp file type, ftp host server, domain name sanger.ac.uk, following the path from "pub" to "databases" to "wormpep" using manual search strings in MSWord 6, identified a number of *C. elegans* ORFs which contained presumptive histidine boxes. Wormpep contains predicted proteins from the *Caenorhabditis elgans* genome sequence project, which is carried out jointly by the Sanger Centre in Cambridge, UK and Genome Sequencing Center in St. Louis, USA. The current Wormpep database, Wormpep 16, contains 16,332 protein sequences (7,120,115 residues). Search strings used included [HXXHH], [HXXXHH], [QXXHH] and [YHH]. Comparison of the data from the two different searches indicated a small (<10) number of putative ORFs as candidate elongases. The histidine box motifs are located at amino acids 162-166 of SEQ ID NO:15, amino acids 186-190 of SEQ ID NO:16, amino acids 145-150 of SEQ ID NO:17, amino acids 147-151 of SEQ ID NO:18, amino acids 141-145 of SEQ ID NO:19, amino acids 177-181 of SEQ ID NO:20, amino acids 155-159 of SEQ ID NO:21, and amino acids 233-237 of SEQ ID NO:22.

(7) On page 10, delete the paragraph beginning on line 7 and substitute the following paragraph:

Since the inventors had previously observed that *C. elegans* genes involved in the synthesis of PUFA may exist in tandem (for example the Δ5 and Δ6 desaturases required for AA and GLA synthesis, respectively, are <1 kB apart on chromosome IV (Michaelson et al., (1998), *FEBS Letts* **439**, 215-218), the positions of the putative *C. elegans* elongase ORFs were determined

using the Sanger Centre's WebAce *C. elegans* server identified with the URL address http filetype, www host server, domain name sanger.ac.uk and following the path from "Projects" to C_elegans" to "webac_front_ends.shtml". This indicated that two pairs of putative elongases were in close proximity to each other on the *C. elegans* chromosome IV.

(8) On page 13, delete the paragraph beginning at line 2 and substitute the following paragraph:

Putative elongase sequences F56H11.4 and F41H10.8 were cloned by PCR into the pYES2 vector (Invitrogen). A *C. elegans* mixed stage cDNA library was used as a PCR template.
F56H11.4 was amplified using primers:
56h114.for 5'-GCGGGTACCATGGCTCAGCATCCGCTC-3' (SEQ ID NO:1) and;
56h114.rev 5'-GCGGGATCCTTAGTTGTTCTTCTTCTT-3' (SEQ ID NO:2).
F41H10.8 was amplified using primers:
41h108.for 5'-GCGGGTACCATGCCACAGGGAGAAGTC-3' (SEQ ID NO:3) and;
416h108.rev 5'-GCGGGATCCTTATTCAATTTTTCTTTT-3' (SEQ ID NO:4).

(9) On page 13, delete the paragraph beginning on line 13 and substitute the following paragraph:

An ORF encoding the *Mortierella alpina* Δ^5-fatty acid desaturase (Michaelson, L.V., et al. (1998) *J. Biol. Chem.*, **273**, 19055-19059) was amplified using primers:
Mad5.for 5'-GCGAATTCACCATGGGTACGGACCAAGGA-3' (SEQ ID NO:5) and;
Mad5.rev 5'-GCGGAGCTCCTACTCTTCCTTGGGACG-3' (SEQ ID NO:6).

(10) Delete the informal sequence listing at pages 22-27 and insert the paper copy of the formal sequence listing at the end of the application.